EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Faculty of Philosophy
General Linguistics

Semantics & Pragmatics SoSe 2022

Lecture 12: Current Research and Applications

14/06/2022, Christian Bentz

Overview

Q&A Tutorial 5

In Exercise 1, why does it say "x, y, z, etc. (in alphabetical order), and then with X, Y, Z, etc. (in alphabetical order)"? Are these rules of lambda-calculus?

No, these are not strict rules of lambda-abstraction. For an expression Z(x), for instance, you could also first abstract over Z and then over x to yield $\lambda x(\lambda Z(Z(x)))$, rather than the other way around, i.e. $\lambda Z(\lambda x(Z(x)))$. I just introduced these additional conventions to have a single solution, rather than multiple possible solutions.

However, note that the order does indeed make a difference for the *types* of the lambda expressions: namely, the former is of type $\langle e, \langle\langle e, t\rangle, t\rangle\rangle$, while the latter is of type $\langle\langle e, t\rangle, \langle e, t\rangle\rangle$.

Q&A Tutorial 5

In Exercise 3, given the sentence "Jumbo hits the tree" why can we not just represent "the tree" as D(T) and "the" as D, rather than using the more complicated λ-expressions?

Actually, you can. In fact, D(T) and D are also valid expressions in the type-theoretic language as we have defined it in the lecture. Clause (i) of the syntax permits any variable and constant of any arbitrary type to be a valid expression, and clause (ii) allows for functional application of these given they have fitting types. So H(D(T)), D(T) and D are valid expressions too.

However, I requested you to use λ-expressions, for the reason that in other sentence constructions you would have to use them. Imagine, for instance, the sentence *Jumbo hits the tree and the rock.* If you want to represent just *hits the tree and the rock* now, H(D(T)) $\wedge$ H(D(R)) will not do, since the expressions connected by logical "and" are both of type $\langle e, t\rangle$, while they would have to be of type t according to clause (iii) of the type-theoretic syntax. You can overcome this problem by using the λ-expression λx(H(D(T))(x) $\wedge$ H(D(R))(x)). Now, both expressions combined with the logical operator are of type t, the overall expression is of type $\langle e, t\rangle$.

Faculty of Philosophy
General Linguistics

Section 1: Information Theory

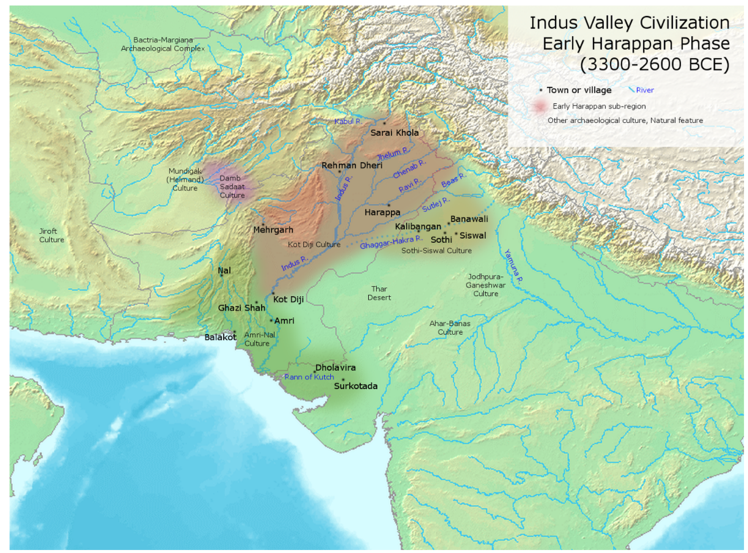
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Faculty of Philosophy
General Linguistics

Section 1.1: Entropic Investigations of Undeciphered Scripts

Entropic Analyses of Undeciphered Scripts

(a) (b)



Rao et al. (2009). Entropic evidence for linguistic structure in the Indus script.
Rao (2010). Probabilistic analysis of an ancient undeciphered script.
Rao et al. (2010). Entropy, the Indus script, and language.

Entropic Analyses: Block Entropy

Q&As
Section 1: Information Theory
Section 2: Formal Semantics
References

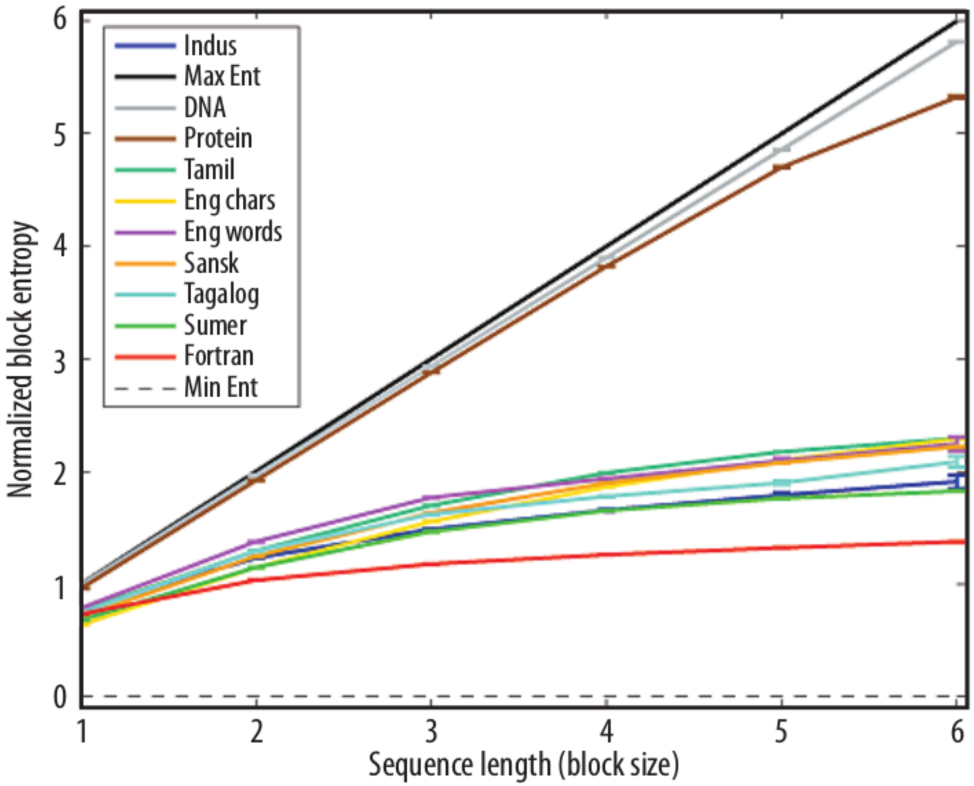
“Block entropy for block size N is defined as:

$$H_N = -\sum_i p_i^{(N)} \log p_i^{(N)} \tag{1}$$

where $p_i^{(N)}$ are the probabilities of sequences (blocks) of N symbols. Thus for $N = 1$, block entropy is simply the standard unigram entropy and for $N = 2$, it is the entropy of bigrams.”

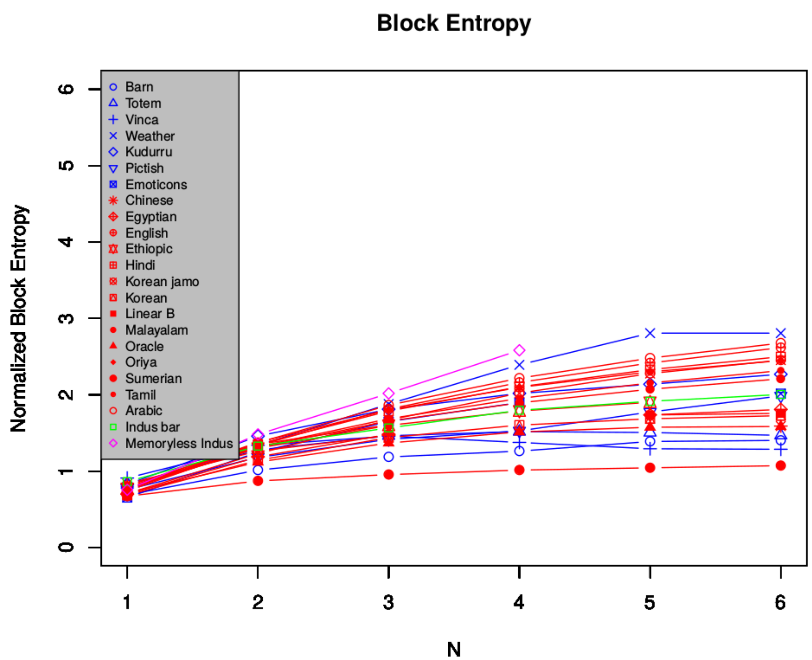
Rao et al. (2010). Entropy, the Indus script, and language, p. 4

Written language or not?



Rao (2010). Probabilistic analysis of an ancient undeciphered script.

However...



"Using a larger set of nonlinguistic and comparison linguistic corpora than were used in these and other studies, I show that none of the previously proposed methods are useful as published. However, one of the measures proposed by Lee and colleagues (2010a) (with a different cut-off value) and a novel measure based on repetition turn out to be good measures for classifying symbol systems into the two categories."

Sproat (2014). A statistical comparison of written language and nonlinguistic symbol systems.

Summary

- A series of studies proposed to use **entropic measures** to distinguish human writing from other types of symbol systems.
- However, the usefulness of these measures has been called into question and needs further investigation.

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Faculty of Philosophy
General Linguistics

Section 1.2: Measuring Morphological Complexity

Measuring Morphological Complexity

Languages differ with regards to how productively they apply bound morhphemes to encode information about gender, case, tense, etc. How can we measure such differences?

	SINGULAR				PLURAL			
CLASS	NOM	GEN	ACC	VOC	NOM	GEN	ACC	VOC
1	-os	-u	-on	-e	-i	-on	-us	-i
2	-s	-Ø	-Ø	-Ø	-es	-on	-es	-es
3	-Ø	-s	-Ø	-Ø	-es	-on	-es	-es
4	-Ø	-s	-Ø	-Ø	-is	-on	-is	-is
5	-o	-u	-o	-o	-a	-on	-a	-a
6	-Ø	-u	-Ø	-Ø	-a	-on	-a	-a
7	-os	-us	-os	-os	-i	-on	-i	-i
8	-Ø	-os	-Ø	-Ø	-a	-on	-a	-a

TABLE 1. Modern Greek nominal inflection classes (Ralli 1994, 2002).

Ackerman & Malouf (2013). Morphological organization: The low conditional entropy conjecture.

Measuring Morphological Complexity

Languages differ with regards to how productively they apply bound morhphemes to encode information about gender, case, tense, etc. How can we measure such differences?

	SINGULAR	PLURAL
CLASS	NOM	NOM
1	-∅	-s
2	-∅	-∅
3	-∅	-en

Table 2: Modern English nominal inflection classes for comparison. I have left out irregular nouns with ablaut (e.g. *man/men*, *foot/feet*), as well as foreign loanwords (*criterion/criteria*). Also, genitive *'s* is not considered an inflectional affix but rather a clitic.

Enumerative Complexity

"**Enumerative complexity (E-complexity)** reflects the number of morphosyntactic distinctions that languages make and the strategies employed to encode them, [...]"

Integrative Complexity

"The **I-complexity** of an inflectional system reflects the difficulty that a paradigmatic system poses for language users (rather than lexicographers) in information-theoretic terms."

Ackerman & Malouf (2013). Morphological organization: The low conditional entropy conjecture.

Average Entropy (E-Complexity)

"The average entropy of a paradigm is the **uncertainty in guessing the realization for a particular cell of the paradigm of a particular lexeme** (given knowledge of the possible exponents). This gives one a measure of the complexity of a morphological system – systems with more exponents and more inflection classes will in general have higher average paradigm entropy [...]"

Average Conditional Entropy (I-Complexity)

[...] Thus, a better **measure of morphological complexity** is the average **conditional entropy**, the average uncertainty in guessing the realization of one randomly selected cell in the paradigm of a lexeme given the realization of one other randomly selected cell. This is the **I-complexity** of paradigm organization."

Ackerman & Malouf (2013). Morphological organization: The low conditional entropy conjecture.

Paradigm Cell Entropy

We can calculate the entropy for every declension, i.e. paradigm cell (corresponding to columns in the table below), across the different classes and their morphological markers. This is called the **Paradigm Cell Entropy (H(c))**.

	SINGULAR				PLURAL			
CLASS	NOM	GEN	ACC	VOC	NOM	GEN	ACC	VOC
1	-os	-u	-on	-e	-i	-on	-us	-i
2	-s	-Ø	-Ø	-Ø	-es	-on	-es	-es
3	-Ø	-s	-Ø	-Ø	-es	-on	-es	-es
4	-Ø	-s	-Ø	-Ø	-is	-on	-is	-is
5	-o	-u	-o	-o	-a	-on	-a	-a
6	-Ø	-u	-Ø	-Ø	-a	-on	-a	-a
7	-os	-us	-os	-os	-i	-on	-i	-i
8	-Ø	-os	-Ø	-Ø	-a	-on	-a	-a

TABLE 1. Modern Greek nominal inflection classes (Ralli 1994, 2002).

Example: Nominative Singular

The entropy is defined as:

$$H(X) = -\sum_i p(x_i) \log_2 p(x_i), \tag{2}$$

where each $p(x_i)$ is the relative frequency of each marker in the cell (i.e. maximum likelihood estimator). For instance, $p(\text{-os}) = \frac{2}{8}$ for the nominative singular cell.

The overall entropy of the nominative singular cell is thus:

$$-(\frac{2}{8} \times \log_2(\frac{2}{8}) + \frac{1}{8} \times \log_2(\frac{1}{8}) + \frac{4}{8} \times \log_2(\frac{4}{8}) + \frac{1}{8} \times \log_2(\frac{1}{8})) = \mathbf{1.75} \text{ bits/inflection} \tag{3}$$

CLASS	NOM
1	-os
2	-s
3	-Ø
4	-Ø
5	-o
6	-Ø
7	-os
8	-Ø

Average Paradigm Entropy (E-Complexity)

The average of Paradigm Cell Entropies across all cells (columns) is then the **Average Paradigm Entropy**.

	SINGULAR				PLURAL			
CLASS	NOM	GEN	ACC	VOC	NOM	GEN	ACC	VOC
1	-os	-u	-on	-e	-i	-on	-us	-i
2	-s	-Ø	-Ø	-Ø	-es	-on	-es	-es
3	-Ø	-s	-Ø	-Ø	-es	-on	-es	-es
4	-Ø	-s	-Ø	-Ø	-is	-on	-is	-is
5	-o	-u	-o	-o	-a	-on	-a	-a
6	-Ø	-u	-Ø	-Ø	-a	-on	-a	-a
7	-os	-us	-os	-os	-i	-on	-i	-i
8	-Ø	-os	-Ø	-Ø	-a	-on	-a	-a

TABLE 1. Modern Greek nominal inflection classes (Ralli 1994, 2002).

(8)

c	NOM.SG	GEN.SG	ACC.SG	VOC.SG	NOM.PL	GEN.PL	ACC.PL	VOC.PL
$H(c)$	1.750	2.156	1.549	1.549	1.906	0.000	2.156	1.906

AVG
1.621

Average Paradigm Entropy (AVG ENTROPY) Across 10 Languages

LANGUAGE	CELLS	REALIZATIONS	MAX REALIZATIONS	DECL CLASSES	DECL ENTROPY	PARADIGM ENTROPY	AVG ENTROPY
Amele	3	30	14	24	4.585	1.105	2.882
Arapesh	2	41	26	26	4.700	0.630	4.071
Burmeso	12	6	2	2	1.000	0.000	1.000
Fur	12	50	10	19	4.248	0.517	2.395
Greek	8	12	5	8	3.000	0.644	1.621
Kwerba	12	9	4	4	2.000	0.428	0.864
Mazatec	6	356	94	109	6.768	0.709	4.920
Ngiti	16	7	5	10	3.322	0.484	1.937
Nuer	6	3	2	16	4.000	0.750	0.778
Russian	12	14	3	4	2.000	0.538	0.911

TABLE 3. Paradigm entropies.

Ackerman & Malouf (2013). Morphological organization: The low conditional entropy conjecture.

Example: Burmeso (bzu, Isolate, Papunesia)

The Average Paradigm Entropy of Burmeso is **relatively low**, namely, exactly **1 bit/inflection**. Note that this is because in the paradigm used (table below) there are two inflectional classes with always two different inflectional markers, so there is consistently 1 bit of choice.

CLASS	I.SG	I.PL	II.SG	II.PL	III.SG	III.PL	IV.SG	IV.PL	V.SG	V.PL	VI.SG	VI.PL
A	j-	s-	g-	s-	g-	j-	j-	j-	j-	g-	g-	g-
B	b-	t-	n-	t-	n-	b-	b-	b-	b-	n-	n-	n-

TABLE A3. Burmeso object agreement prefixes (Donohue 2001, Baerman et al. 2010).

Ackerman & Malouf (2013). Morphological organization: The low conditional entropy conjecture.

Example: Chiquihuitlán Mazatec (maq, Otomanguean, North America)

The Average Paradigm Entropy of Chiquihuitlán Mazatec is **relatively high**, namely, **4.9 bit/inflection**.

CLASS	1SG	2SG	3	1INCL	1PL	2PL
1	a/1-1/tsi	e/2-2/nĩ	a/(2-2)/tsi	ã/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
2	a/3-1/tsi	e/3-1/nĩ	a/(3-1)/tsi	ã/3-31/nĩ	ĩ/3-14/nĩ	ũ/3-1/nĩ
3	æ/1-1/tsi	i/2-2/nĩ	i/(2-2)/tsi	ẽ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
4	u/1-1/tsi	i/2-2/nĩ	u/(2-2)/tsi	ũ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
5	a/1-1/bi	e/2-2/bi	a/(2-2)/bi	ã/2-2/bi	ĩ/2-24/bi	ũ/2-2/bi
6	a/3-1/bi	e/2-2/bi	a/(3-2)/bi	ã/2-2/bi	ĩ/2-24/bi	ũ/2-2/bi
7	u/3-1/tsi	i/3-1/nĩ	u/(3-1)/tsi	ũ/3-31/nĩ	ĩ/3-14/nĩ	ũ/3-1/nĩ
8	æ/1-1/tsi	e/2-2/nĩ	e/(2-2)/tsi	ẽ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
9	o/3-1/tsi	e/2-2/nĩ	o/(3-2)/tsi	õ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
10	ẽ/1-43/bi	ĩ/1-43/bi	ẽ/(3-24)/bi	ẽ/14-42/bi	ĩ/14-34/bi	ũ/14-3/bi
11	æ/3-1/tsi	i/3-1/nĩ	i/(3-1)/tsi	ẽ/3-31/nĩ	ĩ/3-14/nĩ	ũ/3-1/nĩ
12	æ/1-1/tsi	e/2-2/nĩ	æ/(2-2)/tsi	ẽ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
13	a/1-43/tsi	e/1-43/nĩ	a/(3-24)/tsi	ã/14-42/nĩ	ĩ/14-34/nĩ	ũ/14-3/nĩ
14	ẽ/1-43/tsi	ĩ/1-43/nĩ	ẽ/(3-24)/tsi	ẽ/14-42/nĩ	ĩ/14-34/nĩ	ũ/14-3/nĩ
15	u/1-43/tsi	i/1-43/nĩ	u/(3-24)/tsi	ũ/14-42/nĩ	ĩ/14-34/nĩ	ũ/14-3/nĩ
16	ẽ/3-1/bu	ĩ/3-1/ču	ẽ/(3-1)/bu	ẽ/3-31/ču	ĩ/3-14/ču	ũ/3-1/ču
17	æ/3-1/tsi	e/3-1/nĩ	æ/(3-1)/tsi	ẽ/3-31/nĩ	ĩ/3-14/nĩ	ũ/3-1/nĩ
18	ũ/1-1/tsi	ĩ/2-2/nĩ	ũ/(2-2)/tsi	ũ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
19	æ/3-1/bi	e/3-1/bi	e/(3-1)/bi	ẽ/3-31/bi	ĩ/3-14/bi	ũ/3-1/bi
20	a/1-1/be	e/2-2/be	a/(2-2)/be	ã/2-2/be	ĩ/2-24/be	ũ/2-2/be
21	u/1-43/be	i/1-43/be	u/(3-24)/be	ũ/14-42/be	ĩ/14-34/be	ũ/14-3/be
22	a/1-1/ba	e/2-2/ba	a/(2-2)/ba	ã/2-2/ba	ĩ/2-24/ba	ũ/2-2/ba
23	a/3-1/ba	e/2-2/ba	a/(3-2)/ba	ã/2-2/ba	ĩ/2-24/ba	ũ/2-2/ba
24	a/1-1/bu	e/2-2/ču	a/(2-2)/bu	ã/2-2/ču	ĩ/2-24/ču	ũ/2-2/ču
25	ẽ/1-43/bu	ĩ/1-43/ču	ẽ/(3-24)/bu	ẽ/14-42/ču	ĩ/14-34/ču	ũ/14-3/ču
26	a/3-1/hba	e/2-2/hba	a/(3-2)/hba	ã/2-2/hba	ĩ/2-24/hba	ũ/2-2/hba
27	æ/3-1/hba	i/2-2/hba	i/(3-2)/hba	ẽ/2-2/hba	ĩ/2-24/hba	ũ/2-2/hba
28	ẽ/1-1/tsi	ĩ/2-2/nĩ	ẽ/(2-2)/tsi	ẽ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
29	æ/3-1/tsi	e/2-2/nĩ	æ/(3-2)/tsi	ẽ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
30	ẽ/3-1/tsi	ĩ/2-2/nĩ	ĩ/(3-2)/tsi	ẽ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
31	æ/3-1/tsi	i/3-1/nĩ	i/(3-1)/tsi	ẽ/3-31/nĩ	ĩ/3-14/nĩ	ũ/3-1/nĩ
32	æ/1-43/tsi	i/1-43/nĩ	i/(3-24)/tsi	ẽ/14-42/nĩ	ĩ/14-34/nĩ	ũ/14-3/nĩ
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	ũ/3-2/nĩ
34	a/3-1/ba	e/3-1/ča	a/(3-1)/ba	ã/3-31/ča	ĩ/3-14/ča	ũ/3-1/ča
35	æ/3-1/ba	e/2-2/ča	e/(3-2)/ba	ẽ/2-2/ča	ĩ/2-24/ča	ũ/2-2/ča

Conditional Entropy

"To quantify the predictability of one form given the other, we can measure the size of the surprise associated with these forms using conditional entropy H(Y|X), the uncertainty in the value of Y given that we already know the value of X."

The conditional entropy for a cell c_1 given a cell c_2 is then defined as:

$$H(c_1|c_2) = \sum_{r_1} \sum_{r_2} P_{c_1}(r_1) P_{c_2}(r_2) \log_2 P_{c_1}(r_1|c_2 = r_2), \qquad (4)$$

where "realizations" (r_1, r_2) stand in for particular inflections.

Example: Genitive Singular and Accusative Plural

Assume we know the ACC.PL inflection is *-i*, then we know that the inflection class is 7, and hence we know that the GEN.SG has to be *-us*. Thus, we have

$$H(\text{GEN.SG}|\text{ACC.PL} = \text{-i}) = 0\,\text{bits}. \tag{5}$$

Assume, on the other hand, we know that the ACC.PL is in *-a*. Inflection classes 5, 6, and 8 are now possible. Hence, it is possible that GEN.SG is either *-u* or *-os*. According to Ackerman & Malouf (2013), p. 441 we then have

$$H(\text{GEN.SG}|\text{ACC.PL} = \text{-a}) = -(\frac{2}{3} \times \log_2 \frac{2}{3} + \frac{1}{3} \times \log_2 \frac{1}{3}) = 0.918\,\text{bits}. \tag{6}$$

	SINGULAR				PLURAL			
CLASS	NOM	GEN	ACC	VOC	NOM	GEN	ACC	VOC
1	-os	-u	-on	-e	-i	-on	-us	-i
2	-s	-Ø	-Ø	-Ø	-es	-on	-es	-es
3	-Ø	-s	-Ø	-Ø	-es	-on	-es	-es
4	-Ø	-s	-Ø	-Ø	-is	-on	-is	-is
5	-o	-u	-o	-o	-a	-on	-a	-a
6	-Ø	-u	-Ø	-Ø	-a	-on	-a	-a
7	-os	-us	-os	-os	-i	-on	-i	-i
8	-Ø	-os	-Ø	-Ø	-a	-on	-a	-a

TABLE 1. Modern Greek nominal inflection classes (Ralli 1994, 2002).

Conditional Paradigm Entropy (I-Complexity)

We can calculate all possible combinations of inflections and their conditional entropies, and then average across them to get the **Average Conditional Paradigm Entropy** (H(P)). This is **0.644** bits for Modern Greek (see lower right corner in table below).

$H(col\|row)$	NOM.SG	GEN.SG	ACC.SG	VOC.SG	NOM.PL	GEN.PL	ACC.PL	VOC.PL	E[row]
NOM.SG	—	1.000	0.250	0.250	0.750	0.000	1.000	0.750	0.571
GEN.SG	0.594	—	0.594	0.594	0.594	0.000	0.594	0.594	0.509
ACC.SG	0.451	1.201	—	0.000	0.951	0.000	0.951	0.951	0.644
VOC.SG	0.451	1.201	0.000	—	0.951	0.000	0.951	0.951	0.644
NOM.PL	0.594	0.844	0.594	0.591	—	0.000	0.250	0.000	0.411
GEN.PL	1.750	2.156	1.549	1.549	1.906	—	2.156	1.906	1.853
ACC.PL	0.594	0.594	0.344	0.344	0.000	0.000	—	0.000	0.268
VOC.PL	0.594	0.844	0.594	0.594	0.000	0.000	0.250	—	0.411
E[col]	0.719	1.120	0.561	0.561	0.736	0.000	0.879	0.736	0.664

TABLE 2. Conditional entropies for Modern Greek paradigms in Table 1.

Conditional Paradigm Entropy (PARADIGM ENTROPY) Across 10 Languages

LANGUAGE	CELLS	REALIZATIONS	MAX REALIZATIONS	DECL CLASSES	DECL ENTROPY	PARADIGM ENTROPY	AVG ENTROPY
Amele	3	30	14	24	4.585	1.105	2.882
Arapesh	2	41	26	26	4.700	0.630	4.071
Burmeso	12	6	2	2	1.000	0.000	1.000
Fur	12	50	10	19	4.248	0.517	2.395
Greek	8	12	5	8	3.000	0.644	1.621
Kwerba	12	9	4	4	2.000	0.428	0.864
Mazatec	6	356	94	109	6.768	0.709	4.920
Ngiti	16	7	5	10	3.322	0.484	1.937
Nuer	6	3	2	16	4.000	0.750	0.778
Russian	12	14	3	4	2.000	0.538	0.911

TABLE 3. Paradigm entropies.

Ackerman & Malouf (2013). Morphological organization: The low conditional entropy conjecture.

Are E-complexity and I-complexity related to Learnability?

"[...] a series of experiments [...] confirm that indeed, across a range of paradigms that vary in either e- or i-complexity, neural networks (LSTMs) are sensitive to both, but show a larger effect of e-complexity [...]. In human learners, we fail to find any effect of i-complexity at all."

Johnson et al. (2020b). Predictive structure or paradigm size? Investigating the effects of i-complexity and e-complexity on the learnability of morphological systems.

Johnson et al. (2020a). Assessing integrative complexity as a predictor of morphological learning using neural networks and artificial language learning.

Human Learning

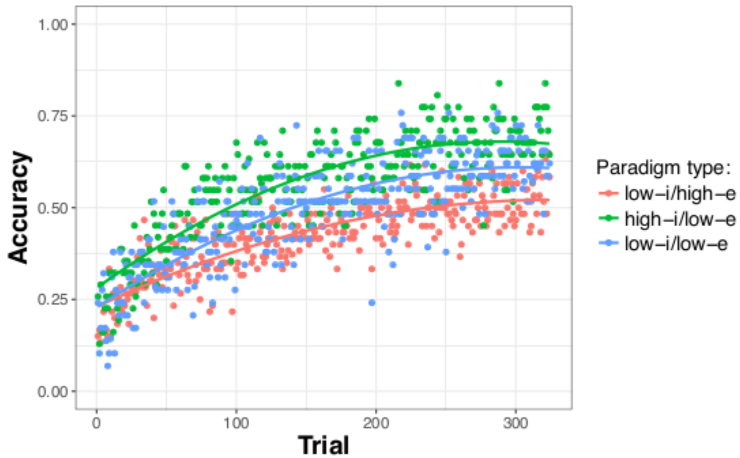
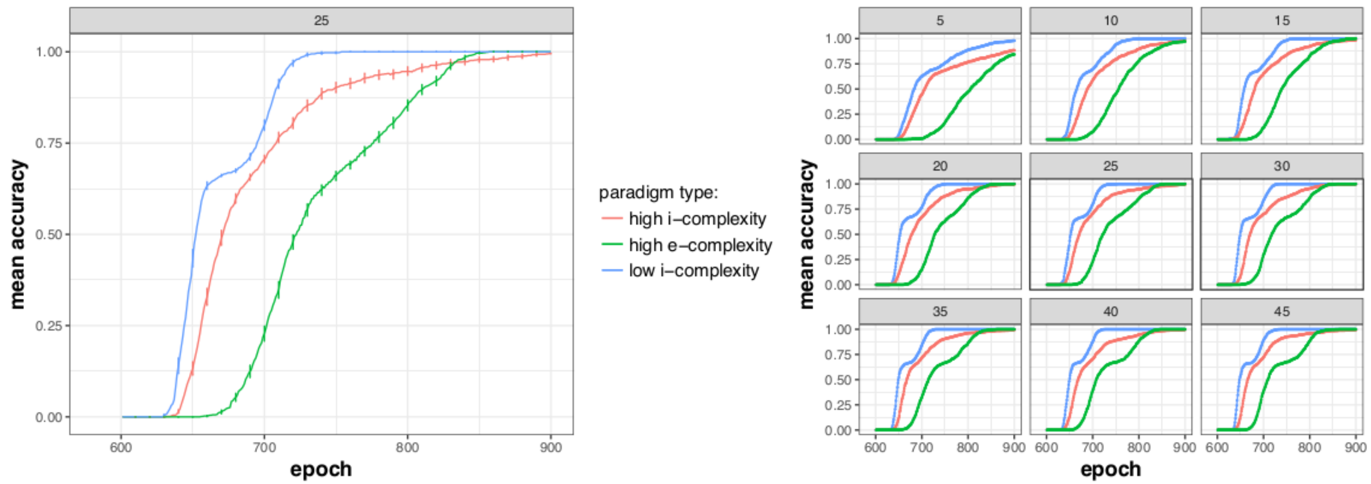


Figure 7. Mean accuracy by trial for each of the three paradigm types (collapsing the two low-i/high-e paradigms). Points indicate the average accuracy across participants for each trial. Lines show quadratic polynomial curves predicting accuracy by trial number for each paradigm type. Learning is worst for the low-i/high-e and best for the high-i/low-e paradigms.

Johnson et al. (2020a).

Machine Learning



A B

Figure 11. Average accuracy across all runs of the LSTM networks in generalizing to novel dual forms for the new high e-complexity paradigm (green) compared with results from the low i-complexity paradigm (blue) and the high i-complexity paradigm (red) from Simulation Experiment 2. A: results for one network size (25 cells), with error bars indicating standard error every 10 epochs. B: results for all the network sizes tested (facet titles give network size in number of cells). Note that the plots start at epoch 600, when the dual forms are introduced to the network (at the beginning of Block 3). In all cases accuracy for the high e-complexity paradigm is lower than for both other paradigms.

Johnson et al. (2020b).

Summary

- Ackerman & Malouf (2013) propose two entropic measures for morphological complexity: the average entropy of a paradigm as a measure of **enumerative complexity**, and the average conditional entropy of cells as an **integrative complexity** measure.
- They argue that the latter is systematically lower (low conditional entropy conjecture), and can be low even for high E-complexity languages.
- They relate I-complexity to **learnability**. However, experiments by Johnson et al. (2020a, 2020b) suggest that E-complexity is more tightly linked with learnability.

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Faculty of Philosophy
General Linguistics

Section 1.3: Syntactic Surprisal

Information Flow in Natural Languages

Q&As
Section 1: Information Theory
Section 2: Formal Semantics
References

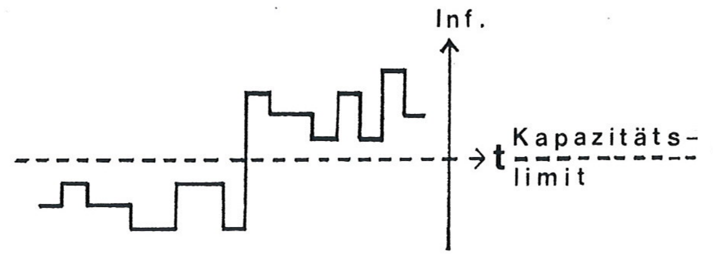


Abb. 1
Schema eines kapazitätsüberfordernden und unökonomischen Informationsflusses

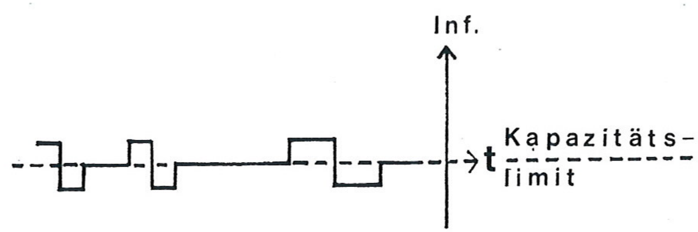


Abb. 2
Schema eines ökonomischen und der Kapazität besser angepaßten Informationsflusses

Note: “Kurzzeitgedächtnis” translates as *short-term memory*.

Fenk & Fenk (1980). Konstanz im Kurzzeitgedächtnis – Konstanz im sprachlichen Informationsfluß?

General Hypothesis

"Soll der Informationsfluß annähernd konstant bleiben, so müssen informationsarme Zeichen bzw. Wörter weniger Zeit – und daher auch weniger Silben [...] – in Anspruch nehmen als informationsreiche [...]"

Fenk & Fenk (1980). Konstanz im Kurzzeitgedächtnis – Konstanz im sprachlichen Informationsfluß?

Translation: In order for the information flow to be approximately constant, it is necessary that signs, i.e. words, with low information content are shorter in time – and hence use fewer syllables [...] – than the ones with high information content [...]

Average Information Content of Mono- and Polysyllabic Words

Berechnung des Informationsgehalts aus den relativen Häufigkeiten verschiedener Wortklassen im Deutschen

x_i	$p(x_i)$	$-\log p(x_i)$	$-p(x_i) \cdot \log p(x_i)$
1-silbig	0,5560	0,2549	0,14173
2-silbig	0,3080	0,5114	0,15752
3-silbig	0,0938	1,0277	0,0964
4-silbig	0,0335	1,4749	0,04941
5-silbig	0,0071	2,1487	0,01525
6-silbig	0,0014	2,8538	0,00399
7-silbig	0,0002	3,6989	0,00073
8-silbig	0,0001	4,0000	0,0004
			0,46543*) dit (= 1,546 bit)

Note: “1-silbig” is *monosyllabic* (e.g. *Baum* “tree”), “2-silbig” is *bisyllabic* (e.g. *On-kel* “un-cle”), etc.

Fenk & Fenk (1980). Konstanz im Kurzzeitgedächtnis – Konstanz im sprachlichen Informationsfluß?

Number of Syllables vs. Information Content

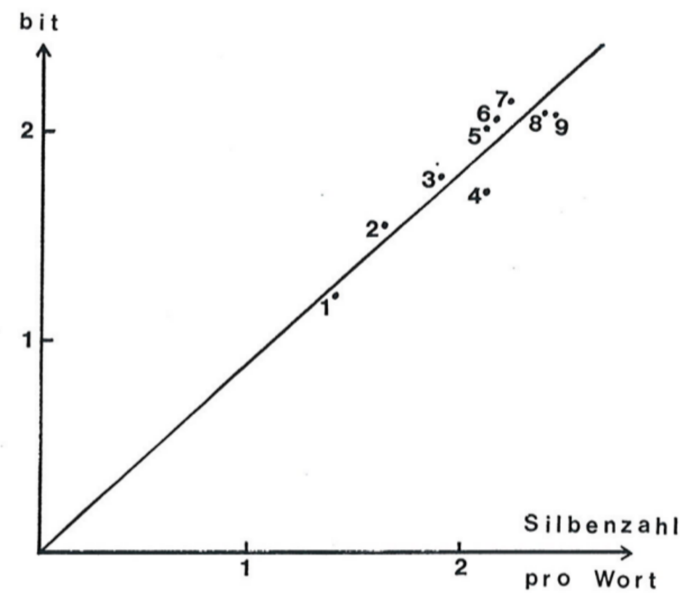


Abb. 3

Verteilung der Sprachen bezüglich mittlerer Wortinformation und mittlerer Wortlänge (1 = Englisch, 2 = Deutsch, 3 = Esperanto, 4 = Arabisch, 5 = Griechisch, 6 = Japanisch, 7 = Russisch, 8 = Latein, 9 = Türkisch)

Note: “Silbenzahl pro Wort” represents average number of syllables per word, and the y-axis (bit) represents the average information content (i.e. entropy).

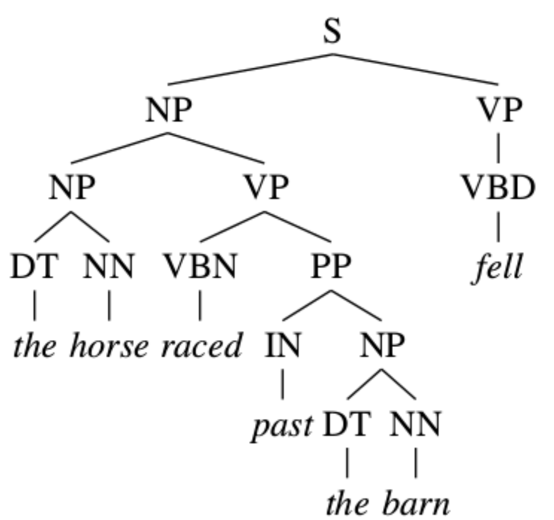
Uniform Information Density Hypothesis (UID)

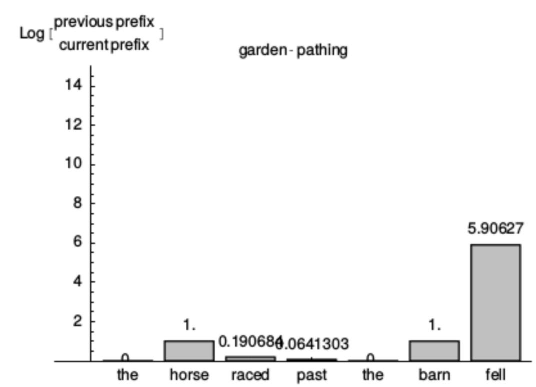
"Short-term-storage and perception mechanisms, which are involved in speech-production and speech-perception, seem to underlie some limitations, that can be defined in terms of information-theory and that should have some effects on languages in the sense of linguistic universals. The hypothesis of **very similar information flow in different languages** could be confirmed [...] in 9 languages."

Fenk & Fenk (1980). Konstanz im Kurzzeitgedächtnis – Konstanz im sprachlichen Informationsfluß?

Syntactic Surprisal: Psycholinguistic Models

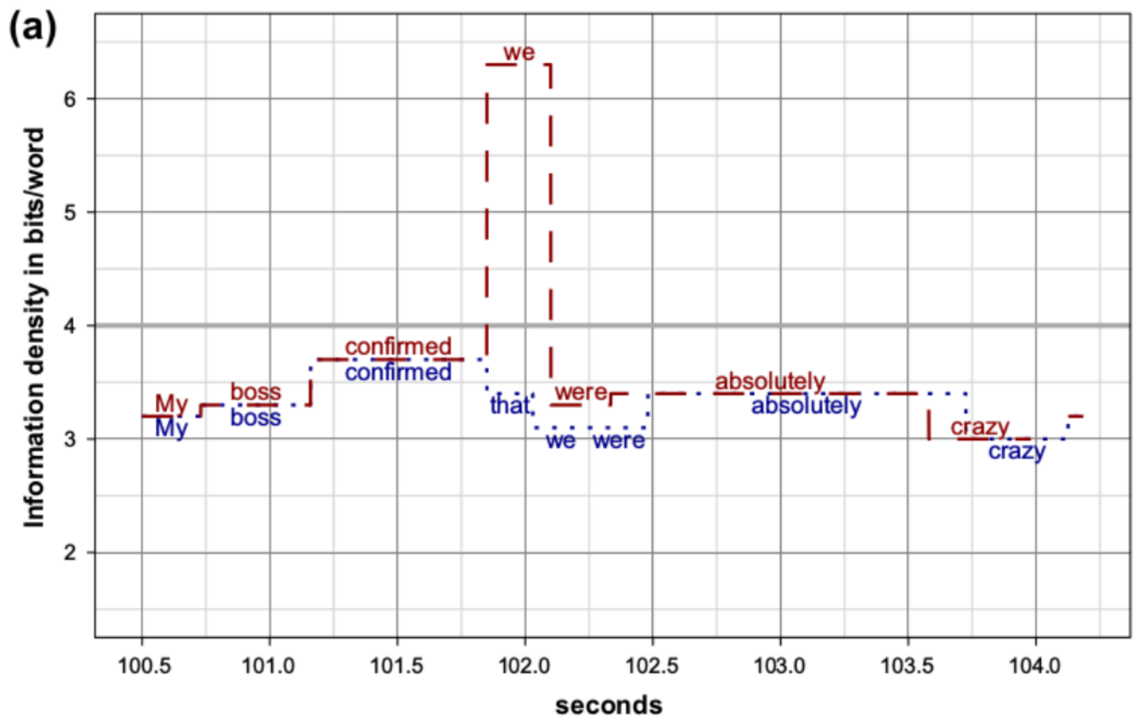
"This report considers a definition of cognitive load in terms of [...] the surprisal of word w_i given its prefix $w_{0...i-1}$ on a phrase-structural language model. [...] Stolcke's probabilistic Earley parser **correctly predicts processing phenomena** associated with garden path structural ambiguity [...]"





Hale (2001). A probabilistic Earley parser as a psycholinguistic model.

Syntactic Surprisal: Usage of Complementizers



Jaeger (2010). Redundancy and reduction: Speakers manage syntactic information density.

Syntactic Surprisal: Gender Paradigms

"These results show that, as expected, in each of the German cases, **gender markers significantly reduce nominal entropy** [...]"

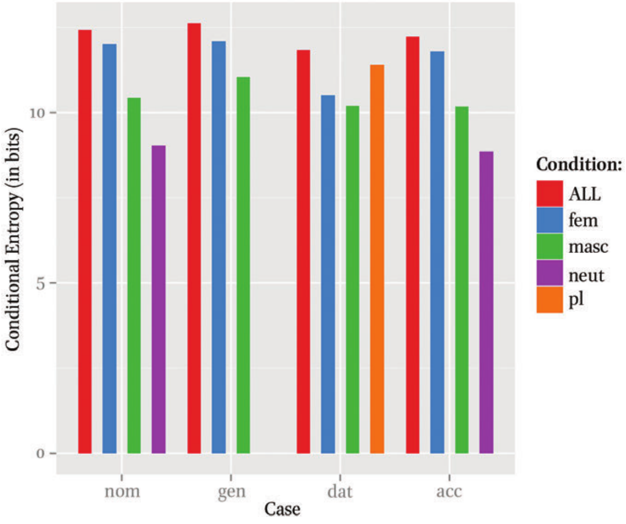
Dye et al. (2017). A functional theory of gender paradigms.

(1)	Gestern	besuchte	ich	**den**	Arzt
	yesterday	visited	I	**the.ACC.MASC**	doctor

"Yesterday I visited the doctor."

"The following noun must belong to the MASCULINE noun class, and thus nouns of all other genders are eliminated as possible candidates in this context. In short, by systematically partitioning nouns into different classes, a gender marker effectively prunes the space of subsequent possibility, delimiting the set of upcoming nouns to class-consistent possibilities."

Figure. Noun entropy conditioned on case and number.



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Faculty of Philosophy
General Linguistics

Section 2: Formal Semantics

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Faculty of Philosophy
General Linguistics

Section 2.1: First Order Logic in NLP

First Order Logic in NLP

"In this paper, we argue that we can combat the data hungriness of neural networks by taking advantage of domain knowledge expressed as **first-order logic**. [...]

While alignments (e.g. *author* to *writing*) can be learned from data, we argue that models can reduce their data dependence if they were guided by easily stated rules such as: Prefer aligning phrases that are marked as similar according to an external resource [...] "

Paragraph: Gaius Julius Caesar (July 100 BC – 15 March 44 BC), Roman general, statesman, Consul and notable author of Latin prose, played a critical role in the events that led to the demise of the Roman Republic and the rise of the Roman Empire through his various military campaigns.

Question: Which Roman general is known for writing prose?

Li & Srikumar (2019). Augmenting neural networks with first-order logic.

First Order Logic in NLP

Predicates

$K_{i,j}$: word p_i is related to q_j in ConceptNet.
$\overleftarrow{A}_{i,j}$: unconstrained model decision that word q_j best matches to p_i.
$\overleftarrow{A}'_{i,j}$: constrained model decision that word q_j best matches to p_i.

FOL rules

R_1: $\forall_{i,j} \in C, K_{i,j} \rightarrow \overleftarrow{A}'_{i,j}$
R_2: $\forall_{i,j} \in C, K_{i,j} \wedge \overleftarrow{A}_{i,j} \rightarrow \overleftarrow{A}'_{i,j}$

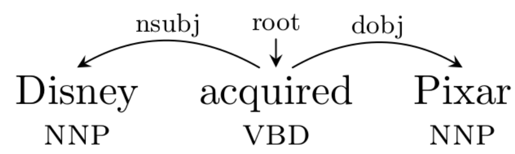
%Train	BiDAF	$+R_1$	$+R_2$	+ELMo	+ELMo,R_1
10%	57.5	**61.5**	60.7	71.8	**73.0**
20%	65.7	**67.2**	66.6	76.9	**77.7**
40%	70.6	**72.6**	71.9	80.3	**80.9**
100%	75.7	**77.4**	77.0	83.9	**84.1**

Li & Srikumar (2019). Augmenting neural networks with first-order logic.

Section 2.1: λ-Calculus in NLP

λ-Calculus in Modern NLP

"We have introduced a method for converting **dependency structures to logical forms using the lambda calculus**. A key idea of this work is the use of a single semantic type for every constituent of the dependency tree, which provides us with a robust way of compositionally deriving logical forms."



(a) The dependency tree for *Disney acquired Pixar*.

$$\text{(nsubj (dobj acquired Pixar) Disney)}$$

(b) The s-expression for the dependency tree.

$$\lambda x.\, \exists yz.\, \mathrm{acquired}(x_e) \wedge \mathrm{Disney}(y_a) \wedge \mathrm{Pixar}(z_a) \\ \wedge \mathrm{arg}_1(x_e, y_a) \wedge \mathrm{arg}_2(x_e, z_a)$$

(c) The composed lambda-calculus expression.

Reddy et al. (2016). Transforming dependency structures to logical forms for semantic parsing.

Neo-Davidsonian Style of Analysis

“We use a version of the lambda calculus with three base types: individuals (**Ind**), events (**Event**), and truth values (**Bool**). Roughly speaking individuals are introduced by nouns, events are introduced by verbs, and whole sentences are functions onto truth values. [...] Verbs such as *acquired* make use of event variables such as x_e, whereas nouns such as *Disney* make use of individual variables such as y_a.”

Reddy et al. (2016)

English words	**Neo-Davidsonian**	**Gamut**
acquired	λx.acquired(x_e)	λx(λy(A(y)(x)))
Disney	λy.Disney(y_a)	λX(X(d))
Pixar	λz.Pixar(z_a)	λX(X(p))

Sentences

Disney acquired Pixar.

λx.$\exists$yz.acquired(x_e) $\wedge$ Disney(y_a) $\wedge$ Pixar(z_a) $\wedge$ arg$_1$(x_e, y_a) $\wedge$ arg$_2$(x_e, z_a)

λx(λy(A(y)(x)))(d)(p)

Performance

"Experiments on the Free917 and Web-Questions datasets show that our representation is superior to the original dependency trees and that it outperforms a CCG-based representation on this task. Compared to prior work, we obtain the strongest result to date on Free917 and competitive results on WebQuestions."

Reddy et al. (2016)

Note: CCG is an abbreviation for *Combinatorial Categorial Grammar*. Free917 is a data set of questions and answers.

Method	Free917 Accuracy	WebQuestions Average F_1
Cai and Yates (2013)	59.0	–
Berant et al. (2013)	62.0	35.7
Kwiatkowski et al. (2013)	68.0	–
Yao and Van Durme (2014)	–	33.0
Berant and Liang (2014)	68.5	39.9
Bao et al. (2014)	–	37.5
Bordes et al. (2014)	–	39.2
Yao (2015)	–	44.3
Yih et al. (2015) (FB API)	–	48.4
Bast and Haussmann (2015)	76.4	49.4
Berant and Liang (2015)	–	49.7
Yih et al. (2015) (Y&C)	–	**52.5**
This Work		
DEPTREE	53.2	40.4
SIMPLEGRAPH	43.7	48.5
CCGGRAPH (+C +E)	73.3	48.6
DEPLAMBDA (+C +E)	**78.0**	50.3

Table 3: Question-answering results on the WebQuestions and Free917 test sets.

Follow-Up Study

"We introduced UDepLambda, a semantic interface for Universal Dependencies, and showed that the resulting semantic representation can be used for question-answering against a knowledge base in multiple languages."

Reddy et al. (2017)

Method	GraphQ.	WebQ.
SEMPRE (Berant et al., 2013)	10.8	35.7
JACANA (Yao and Van Durme, 2014)	5.1	33.0
PARASEMPRE (Berant and Liang, 2014)	12.8	39.9
QA (Yao, 2015)	–	44.3
AQQU (Bast and Haussmann, 2015)	–	49.4
AGENDAIL (Berant and Liang, 2015)	–	49.7
DEPLAMBDA (Reddy et al., 2016)	–	50.3
STAGG (Yih et al., 2015)	–	48.4 (52.5)
BILSTM (Türe and Jojic, 2016)	–	24.9 (52.2)
MCNN (Xu et al., 2016)	–	47.0 (53.3)
AGENDAIL-RANK (Yavuz et al., 2016)	–	51.6 (52.6)
UDEPLAMBDA	17.7	49.5

Table 4: F_1-scores on the English GraphQuestions and WebQuestions test sets (results with additional task-specific resources in parentheses).

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Faculty of Philosophy
General Linguistics

References

References

Dye, M., Milin, P., Futrell, R., & Ramscar, M. (2017). A functional theory of gender paradigms. In F. Kiefer, J. P. Blevins, & H. Bartos (Eds.) *Perspectives on Morphological Organization: Data and Analyses* (pp. 212-239). Brill: Leiden.

Fenk, A. & Fenk, G. (1980). Konstanz im Kurzzeitgedächtnis - Konstanz im sprachlichen Informationsfluß? In: *Zeitschrift für experimentelle und angewandte Psychologie*, Band XXVII, Heft 3, p. 400-414.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies* (pp. 1-8). Association for Computational Linguistics.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive psychology*, 61(1), 23-62.

Johnson et al. (2020a). Assessing integrative complexity as a predictor of morphological learning using neural networks and artificial language learning. *PsyArXiv*.

Johnson et al. (2020b). Predictive structure or paradigm size? Investigating the effects of i-complexity and e-complexity on the learnability of morphological systems. *PsyArXiv*.

Li & Srikumar (2019). Augmenting neural networks with first-order logic. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*.

Reddy, S., Täckström, O., Collins, M., Kwiatkowski, T., Das, D., Steedman, M., & Lapata, M. (2016). Transforming dependency structures to logical forms for semantic parsing. *Transactions of the Association for Computational Linguistics*, 4, 127-140.

Reddy, S., Täckström, O., Petrov, S., Steedman, M., & Lapata, M. (2017). Universal semantic parsing. *arXiv preprint* arXiv:1702.03196.

Rao, R. P., Yadav, N., Vahia, M. N., Joglekar, H., Adhikari, R., & Mahadevan, I. (2009). Entropic evidence for linguistic structure in the Indus script. *Science*, 324(5931), 1165-1165.

Rao, R. P. (2010). Probabilistic analysis of an ancient undeciphered script. *Computer*, (4), 76-80.

Rao, R. P., Yadav, N., Vahia, M. N., Joglekar, H., Adhikari, R., & Mahadevan, I. (2010). Entropy, the Indus script, and language: A reply to R. Sproat. *Computational Linguistics*, 36(4), 795-805.

Sproat, R. (2014). A statistical comparison of written language and nonlinguistic symbol systems. *Language*, 457-481.

Thank You.

Contact:

Faculty of Philosophy
General Linguistics
Dr. Christian Bentz
SFS Wilhelmstraße 19-23, Room 1.24
chris@christianbentz.de
Office hours:
During term: Wednesdays 10-11am
Out of term: arrange via e-mail